Investigating Carbon Trends in the Atmosphere From 1960-Present

Sri-Amirthan Theivendran

Introduction Based on daily measurments of flask CO_2 taken from the Mauna Loa Observatory (MLO) in Hawaii at elevation 3397 m from 1960-present, we investigate several hypotheses. First, although, carbon in the atmosphere is still increasing as shown the graphs below, we will investigate the claim that the increase in carbon has slowed recently. Also we will see whether there is a reasonable chance that carbon exceeds 430 ppm by the year 2025, which month results in most carbon being produced and whether the data supports the slowing of carbon emissions during global recessions and post-collapse of the Soviet Union.

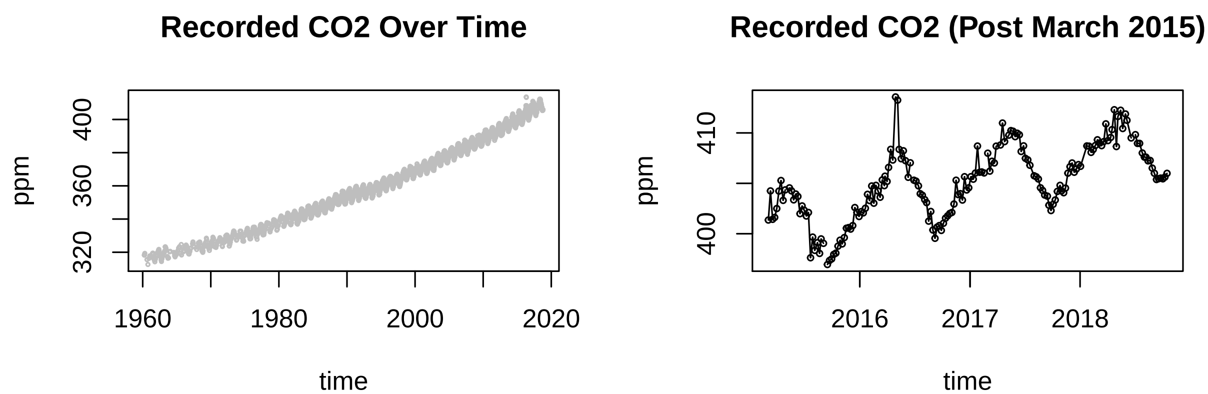

Figure 1: Recorded CO_2

Methods To the above data, we fit a GAM model. Specifically, we assume that the daily measurements $Y_i \sim \text{Gamma}(\mu_i/\nu, \nu)$ where the first parameter is the scale parameter and the second is the shape parameter. Our assumption is based on the following histogram of CO_2 measurements which look roughly gamma distributed. In any case, we use a log-linear model with covariates month, time (relative to 1980), with the log-number of days as an offset, so that $\log \mu_i = X_i'\beta + f(\text{time}_i)$ where X_i is the vector of covariates and f is a cubic spline polynomial. We choose to include month in addition to a smoothly varying time trend in order to account for the apparent seasonality in the data. Obviously, a log-offset is needed to account for the varying numbers of days in months.

Results The parameter estimates for the month coeffients on the natural scale are given below with the baseline being April. We see that carbon tends to higher in March than October. Carbon is highest during Februrary and lowest during October. A graph of the relative rate for the smoothly varying time trend (relative to baseline 1980) along with 95 percent confidence bands is given below. It demonstrates that on the whole carbon in the atmosphere has been increasing for the last 40 years.

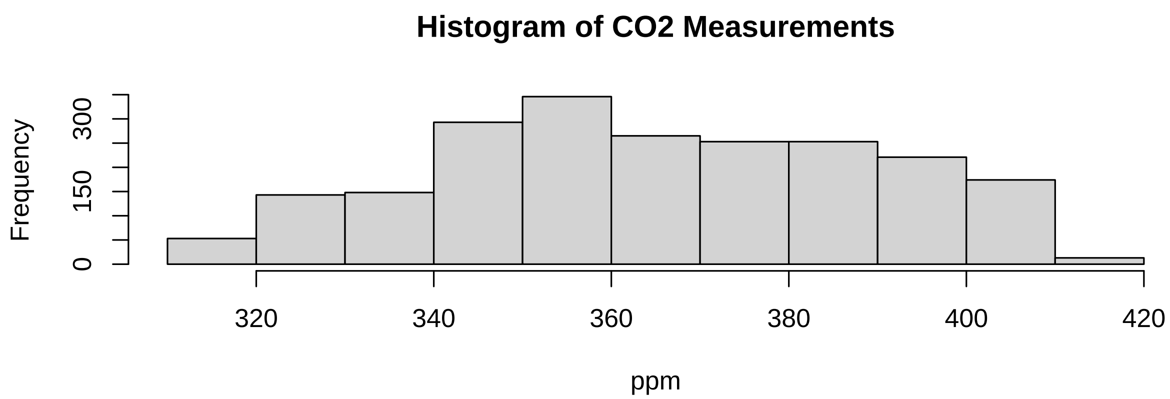

Figure 2: Histogram of CO_2 Measurements

Table 1: Relative rate for Each Month (Baseline is April)

	est	2.5	97.5
monthAugust	0.956	0.955	0.957
monthDecember	0.958	0.957	0.959
monthFebruary	1.057	1.056	1.058
monthJanuary	0.961	0.960	0.962
monthJuly	0.962	0.961	0.962
monthJune	0.999	0.998	1.000
monthMarch	0.965	0.964	0.965
monthMay	0.969	0.968	0.970
monthNovember	0.987	0.986	0.988
monthOctober	0.952	0.951	0.953
monthSeptember	0.983	0.982	0.984

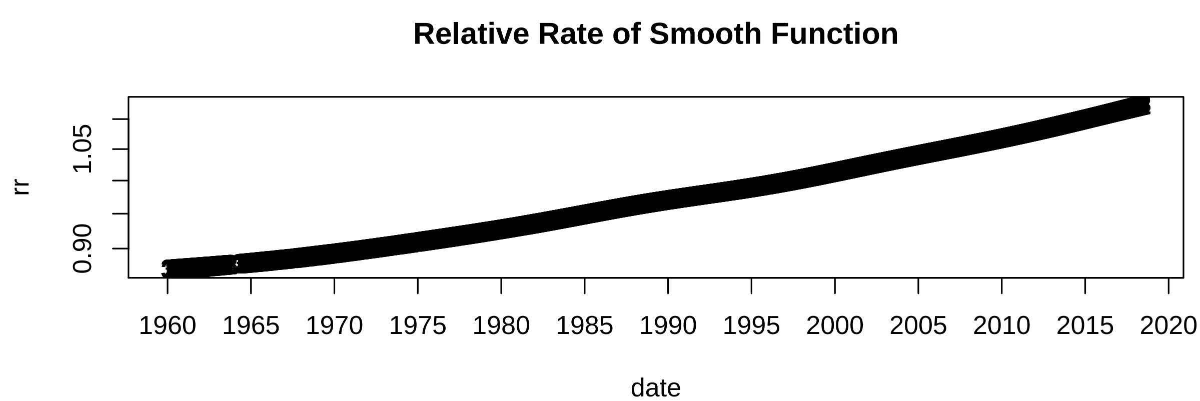

Figure 3: Relative Rate of Smooth Function

Discussion Now we turn our attention to forecasting. Based on the graph of recent CO_2 and predicted CO_2 emissions in the immediate future, we see that carbon will not comfortably exceed 430 ppm until roughly late 2027. Ostenisbly, There is little chance that carbon will exceed 430 ppm by 2025 based on the 95 percent confidence bands. However, we must take this conclusion with a grain of salt as the point confidence intervals glued together do not necessarily construct a confidence interval for the entire function.

It is well-known that carbon levels have been rising in the past 40 years or so (and this observation is consistent with our model) but the rate at which carbon is increasing can fluctuate based on a variety of

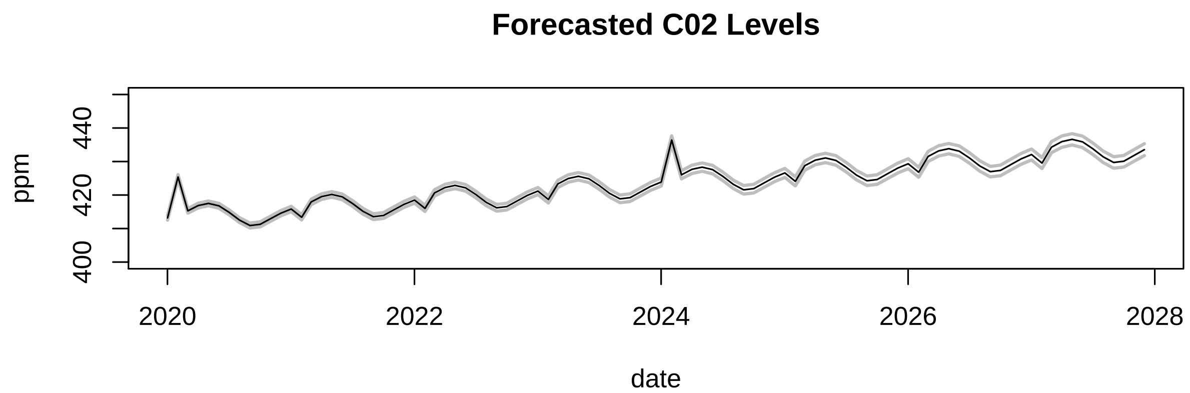

Figure 4: Forecasted Emissions

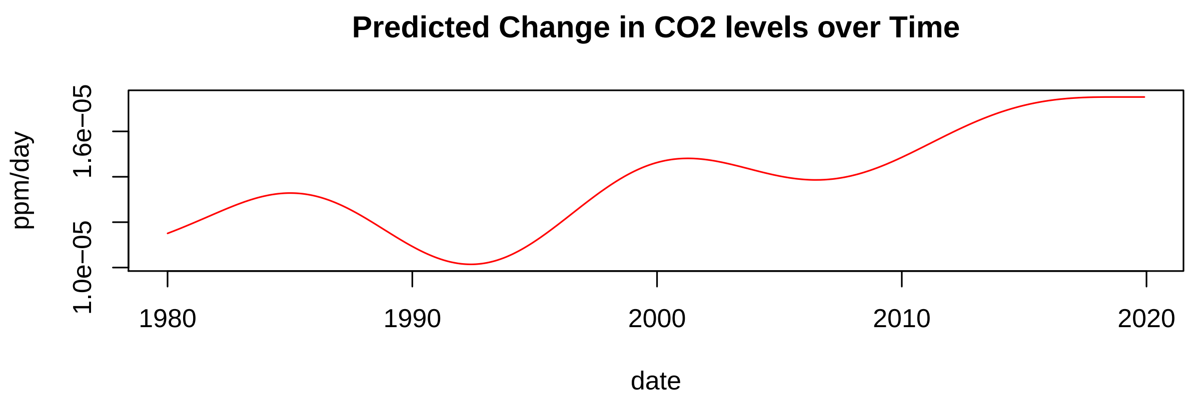

Figure 5: Derivative of Smoothly Varying Function

```r
#setting up the data
cUrl=paste0("http://scrippsco2.ucsd.edu/assets/data/atmospheric/",
            "stations/flask_co2/daily/daily_flask_co2_mlo.csv")
cFile=basename(cUrl)
if(!file.exists(cFile)) download.file(cUrl, cFile)
co2s=read.table(cFile, header=FALSE, sep=",", skip=69,
                stringsAsFactors = FALSE,
                col.names=c("day", "time", "junk1", "junk2",
                            "Nflasks", "quality", "co2"))
co2s$date=strptime(paste(co2s$day, co2s$time), format= "%Y-%m-%d %H:%M", tz="UTC")
co2s[co2s$quality>2, "co2"] = NA

#plot of recent CO2 emissions
par(mfrow=c(1,2))
plot(co2s$date, co2s$co2, cex=0.3,col="grey", xlab="time",
     ylab="ppm", main="Recorded CO2 Over Time")

plot(co2s[co2s$date>ISOdate(2015, 3, 1, tz="UTC"), c("date", "co2")],
     type="o", xlab="time", ylab="ppm", cex=0.5, main="Recorded CO2 (Post March 2015)")
par(mfrow=c(1,1))

#histogram of CO2 measurements
timeOrigin=ISOdate(1980, 1, 1, 0, 0, 0, tz="UTC")
co2s$days=as.numeric(difftime(co2s$date, timeOrigin, units="days"))
co2s$month=as.factor(month.name[as.integer(strftime(co2s$date, "%m"))])
co2s$daysinMonth=Hmisc::monthDays(co2s$date)
co2s$nDays=log(co2s$daysinMonth)
hist(co2s$co2, xlab="ppm", main="Histogram of CO2 Measurements")

#fitting the model and table of coefficients
co2gam=mgcv::gam(co2 ~ month+offset(nDays)+s(days), data=co2s, family = Gamma(link=log))
coefftable=summary(co2gam)$p.table[, 1:2]
paramest=round(exp(coefftable%*%Pmisc::ciMat(0.95))[-1, ], 3)
knitr::kable(paramest, caption="Relative rate for Each Month (Baseline is April)")

#predicted relative rate of the smooth function
co2Pred=as.matrix(as.data.frame(mgcv::predict.gam(co2gam, co2s, type="terms",
                                                  terms="s(days)", se.fit=TRUE)))
co2Pred=exp(co2Pred %*% Pmisc::ciMat())
#confidence bound of the relative rate for the time trend
matplot(co2s$days, co2Pred, log="y", xaxt="n", xlab="date", ylab="rr", col="black",
        lty=c(1,1,2), main="Relative Rate of Smooth Function")
pseq=seq(from=min(co2s$date), by="5 years", length.out = 15)
axis(1, at=difftime(pseq, timeOrigin, units="days"), labels=format(pseq, "%Y") )

#forecasting CO2 emissions from 2020-2028
newX=data.frame(date=seq(from=ISOdate(2020, 1, 1, tz="UTC"), by="months", length.out=12*8))
newX$days=as.numeric(difftime(newX$date, timeOrigin, units="days"))
newX$daysInMonth=Hmisc::monthDays(newX$date)
newX$nDays=log(newX$daysInMonth)
newX$month=as.factor(months(newX$date))
```

```
co2for=predict(co2gam, newX, se.fit=TRUE)
co2for=cbind(newX, co2for)
co2for$lower=co2for$fit-2*co2for$se.fit
co2for$upper=co2for$fit+2*co2for$se.fit
for (D in c("fit", "lower", "upper")){
  co2for[[paste(D, "exp", sep="")]]=exp(co2for[[D]])
}
plot(co2for$date, co2for$fitexp, type="n", xlab="date", ylab="ppm",
     ylim=c(400, 450), main="Forecasted CO2 Levels")
matlines(co2for$date, co2for[, c("lowerexp", "upperexp", "fitexp")],
         lty=1, col=c("grey", "grey", "black"), lwd=c(2,2,1))

#Predicting Derivatives

newX=data.frame(date=seq(from=timeOrigin, by="months", length.out=12*40))
newX$days=as.numeric(difftime(newX$date, timeOrigin, units="days"))
newX$daysInMonth=Hmisc::monthDays(newX$date)
newX$nDays=log(newX$daysInMonth)
newX$month=as.factor(months(newX$date))


#predicting derivatives.
# new data for prediction
# prediction of smoothed estimates at each unique year value
# with standard error
B <- predict(co2gam,  newX, type="response", se.fit=TRUE)


# finite difference approach to derivatives following
# example from ?predict.gam

eps <- 0.5
X0 <- predict(co2gam, newX, type = 'lpmatrix')


newXeps_p <- newX
newXeps_p$days<- newXeps_p$days+eps

X1 <- predict(co2gam, newXeps_p, type = 'lpmatrix')


# finite difference approximation of first derivative
# the design matrix
Xp <- (X1 - X0) / eps

# first derivative
fd_d1 <-  Xp %*% coef(co2gam)
newX$deriv<-fd_d1

#Derivative of  the smoothly varying function
plot(newX$date, newX$deriv, xlab="date", ylab="ppm/day",
     ylim=c(min(newX$deriv), max(newX$deriv)),
     main="Change in CO2 levels over Time", type="l", col="red")
```

Appendix